QASM: a Q&A Social Media System Based on Social Semantics

Zide Meng, Fabien Gandon, Catherine Faron Zucker

QASM: a Q&A Social Media System Based on Social Semantics

Zide Meng[1], Fabien Gandon[1], and Catherine Faron-Zucker[2]

INRIA Sophia Antipolis Méditerranée, 06900 Sophia Antipolis, France[1]
Univ. Nice Sophia Antipolis, CNRS, I3S, UMR 7271, 06900 Sophia Antipolis, France[2]

Abstract. In this paper, we describe the QASM (Question & Answer Social Media) system based on social network analysis to manage the two main resources in CQA sites: users and contents. We first present the QASM vocabulary used to formalize both the level of interest and the expertise of users on topics. Then we present our method to extract this knowledge from CQA sites. Finally we show how this knowledge is used both to find relevant experts for a question and to search for similar questions. We tested QASM on a dataset extracted from the popular CQA site StackOverflow.

Keywords: Community Question Answering, Social Media Mining, Semantic Web

1 Introduction

Community Question Answering (CQA) services provide a platform where users can ask expert for help. Since questions and answers can be viewed and searched afterwards, people with similar questions can also directly find solutions by browsing this content. Therefore, effectively managing these content is a key issue. Previous research works on this topic mainly focus on expert detection [2], similar question retrieval [1]. In this paper, we describe QASM (Question & Answer Social Media), a system based on social network analysis (SNA) to manage the two main resources in CQA sites: users and contents. We first present the QASM vocabulary used to formalize both the level of interest and the expertise of users on topics. Then we present our method to extract this knowledge from CQA sites. Our knowledge model and knowledge extraction method is an extension of our work presented in [3] on social media mining for detecting topics from question tags in CQA sites. Finally we show how this knowledge is used both to find relevant experts for routing questions (users interested and experts in the question topics) and to find answers to questions by browsing CQA content and by identifying relevant answers to similar questions previously posted. We tested QASM on a dataset extracted from the popular CQA site StackOverflow.

2 QASM System Description

2.1 Overview

Figure 1 presents an overview of QASM. We first use the SIOC ontology[1] to construct an RDF dataset from social media data extracted from a CQA site. Then we use social media mining techniques to extract topics, interests and expertise levels from this dataset. We formalize them with the QASM schema and enrich our RDF dataset with this knowledge. As a result, we provide an integrated and enriched Q&A triple store which contains both user interests, levels of expertise and topics learned from question tags. Finally, we linked our dataset with DBpedia (through named entity identification).

Based on the QASM RDF dataset, we can provide the users of the Q&A site with two services to find relevant experts for a question and to search for similar questions. We detail them in the following subsections.

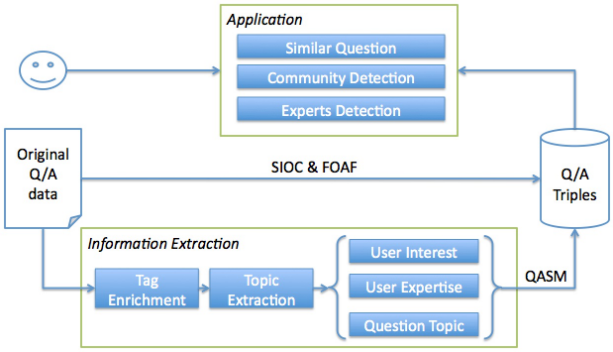

Fig. 1. Overview of QASM

2.2 QASM Vocabulary

The QASM vocabulary[2] enables to model the level of user interests and expertise and topics of questions and answers from Q&A sites. Figure 2 provides an overview of it. It reuses both the SIOC ontology and the Weighting ontology[3].

– `qasm:Topic` represents a set of tags related to a specified topic. In our models, tags belong to instances of `qasm:Topic`, we also consider different tags have different weights for each topic.

[1] http://sioc-project.org/ontology
[2] It is available online at http://ns.inria.fr/qasm/qasm.html
[3] http://smiy.sourceforge.net/wo/spec/weightingontology.html

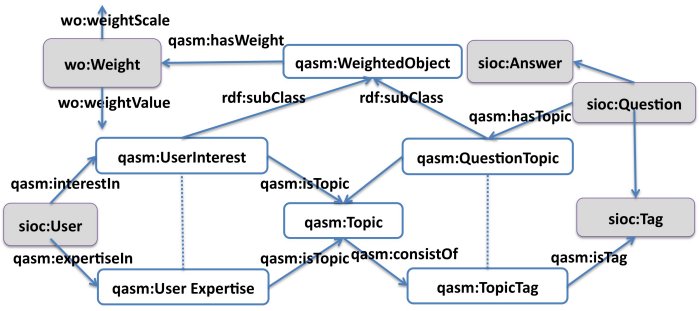

Fig. 2. Overview of the QASM vocabulary

- `qasm:WeightedObject` is used to describe the weight that a specified subject has with regard to a specified object. This class has four subclasses which represent question topics, users' interests, users' expertise and tag topics respectively. In fact, this class is used to model the distributions we extracted from the original data. For example, topic-tag distribution, user-interest distribution.
- `qasm:interestIn` is used to describe the user-interest distribution. This property is different from `foaf:interest` for its range. In FOAF people are interested in documents, while in QASM a user is interested in a topic to a certain degree (a weight).
- `qasm:expertiseIn` is used to describe the user-expertise distribution. A user has different weights for different topics.

2.3 Knowledge Extraction by Social Media Mining

Topics, interests and levels of expertise are implicit information in the available raw CQA data. We use social media mining techniques to extract this knowledge.

- Topics & User Interests In [3], we proposed a light-weight model to extract topics from question tags. The output of this model is a topic-tag distribution where each tag belonging to a topic is given a weight (probability) indicating to what extent the tag is related to the topic. A user answering a question acquires the tags attached to this question and can therefore be represented by a list of tags. Then we use the topic-tag distribution to compute a user-topic distribution indicating to what extent each user is related to a topic.
- User Expertise The users interested in a question may provide answers to it or comments to other answers. Each question or answer may get votes from other users and an answer may be chosen as the best answer. By exploiting the tags attached to a question and the topic-tag distribution, the users providing questions or answers with a high number of votes or the best answers can be considered as experts in the topics to which their questions belongs. Equation 1 defines how we use the vote information to compute users' levels of expertise. $E_{u,k}$ denotes the expertise of user u on topic k, m denotes the number of answers provided by user u, $P_{t,k}$ denotes the weight

of tag t for topic k, Q_i and $A_{i,j}$ denote the votes on question i and its j^{th} answer, where A_j is the j^{th} answer provided by user u to question Q_i.

$$E_{u,k} = \sum_{i=1}^{m} P_{t,k} * log(Q_i) * log(A_{i,j}) \quad (1)$$

2.4 Experimental Evaluation

We first built an RDF dataset from Stackoverflow raw data which comprises 15327727 triples[4]. Then we randomly chose several questions and for each question we recorded 10 or 20 users provided by our system. Then for each question, we computed the proportion of the recorded users who actually answered it. Compared to [4], our results are much better.

Table 1. Preliminary results on question routing

	100	500	1000	average	[4]
precision@10	0.021	0.0188	0.0187	**0.0195**	0.0167
precision@20	0.016	0.0134	0.0134	**0.0143**	0.0118

3 Conclusion and Future Work

We presented QASM, a Q&A system combining social media mining and semantic web models and technologies to manage Q&A users and content in CQA sites. There are many potential future directions for this work. We are currently considering constructing a benchmark for Q&A system based on our Stackoverflow dataset. In a near future we will also enrich the linking of QASM with the LOD which may help to improve question routing and similar question search.

References

1. Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J.: Discovering value from community activity on focused question answering sites: a case study of stack overflow. In Proc. of the 18th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (2012)
2. Yang, L., Qiu, M., Gottipati, S., Zhu, F., Jiang, J., Sun, H., Chen, Z.: CQArank: jointly model topics and expertise in community question answering. In Proc. of the 22nd ACM Int. Conf. on Information & Knowledge Management (2013)
3. Zide, M., Fabien, G., Catherine, F., Ge, S.: Empirical Study on Overlapping Community Detection in Question and Answer Sites. In Proc. of the Int. Conf. on Advances in Social Networks Analysis and Mining (2014)
4. Chang, S., Pal, A.: Routing questions for collaborative answering in community question answering. In Proc. of Int. Conf. on Advances in Social Networks Analysis and Mining (ASONAM) (2013)

[4] It is available online at https://wimmics.inria.fr/data